

# Open Science in the Two!Ears Project Experiences and Best Practices

Hagen Wierstorf<sup>1</sup>   Fiete Winter<sup>2</sup>   **Sascha Spors<sup>2</sup>**

<sup>1</sup>Centre for Vision, Speech and Signal Processing, University of Surrey

<sup>2</sup>Institute of Communications Engineering, University of Rostock

# The Elements of Open Science

Open Source

Open Data

Open Access

Open Methodology

Open Notebook  
Science

Open Educational  
Resources

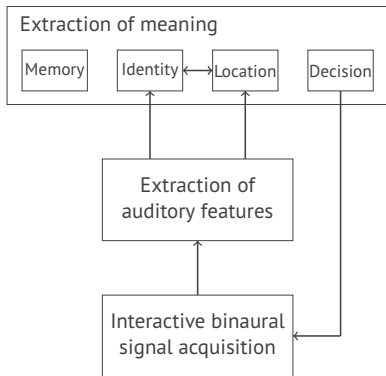
Open Peer Review

Open Research

- Requires systematic management of research data
- Open science by itself does not ensure the ease of reproducibility

# Two!Ears

## Computational framework for modelling active exploratory listening that assigns meaning to auditory scenes



- 9 international partners
  - Scenario-based development and evaluation
  - Modular architecture
  - Simulated/real-world input
- ⇒ **open source software and data**

# Comparison with AMToolbox

## Auditory Modeling Toolbox

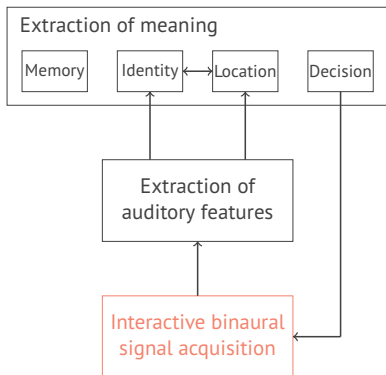
- AABBA project started in 2009: **apply binaural models**
- Source code of models rarely available
- Initiated open collection of models:  
<http://amtoolbox.sourceforge.net>

## Additional features required by Two!Ears

- Block-based processing
- Clear separation in feature extraction and machine learning
- Seamless combination of different approaches
- Initiated dedicated modeling approach:  
<http://docs.twoears.eu>

# Two!Ears Auditory Model

## Binaural Simulator



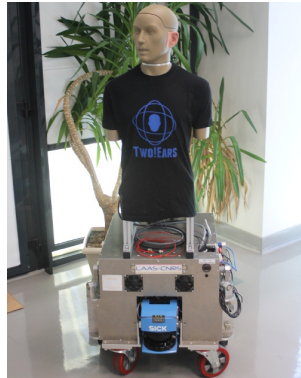
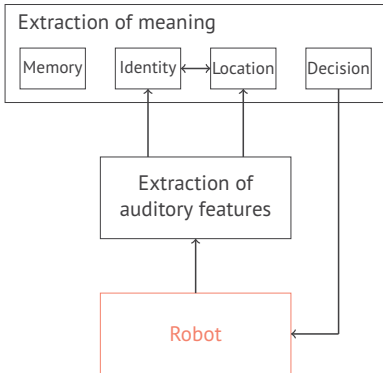
The screenshot shows the GitHub repository for the `Two!Ears / binaural-simulator`. The repository is a module for the `Two!Ears Auditory Model - Binaural simulator`, with the latest commit on Jan 17. The repository statistics include 375 commits, 6 branches, 4 releases, and 6 contributors. The commit history is as follows:

Commit	Revised description	Time ago
<code>doc</code>	remove doxygen directory	2 years ago
<code>src</code>	Merge branch 'master' of https://github.com/Two!EARS/binaural-simulator	7 months ago
<code>test</code>	add methods to incorporate head rotation limits and torso orientations	8 months ago
<code>tmp</code>	add tmp directory with gllignore	3 years ago
<code>gllignore</code>	added '~' to gllignore	8 months ago
<code>BinauralSimulator.xml</code>	Remove dependency on Tools repo	2 years ago
<code>LICENSE</code>	Added GPL2 as license	3 years ago
<code>README.md</code>	Revised description	5 months ago

- Defined interface to other modules
- May be used standalone

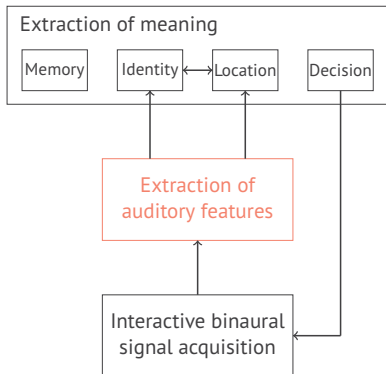
# Two!Ears Auditory Model

Real-World input by robot



# Two!Ears Auditory Model

Auditory front-end



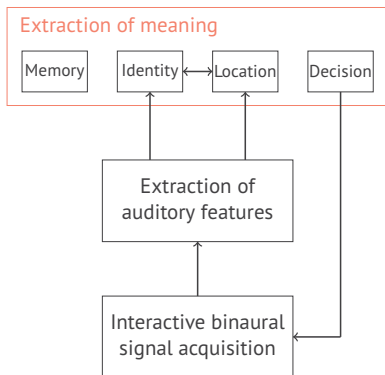
The screenshot shows the GitHub repository page for 'Two!Ears / auditory-front-end'. The repository has 689 commits, 2 branches, 3 releases, and 7 contributors. The commit history is as follows:

Commit	Message	Time
doc	Remove outdated user manual and replace by website	2 years ago
src	Accomodate size() dimension arg (#23)	6 months ago
test	Add precedence model processor (#18)	6 months ago
.gitignore	ignore additional autosaves of matlab editors for different platforms	a year ago
AuditoryFrontEnd.xml	Add precedence model processor (#18)	6 months ago
LICENSE	Add GPL2 license	3 years ago
README.md	Update doc link	6 months ago
startAuditoryFrontEnd.m	Include test folder in the AFE startup script	6 months ago

- Extracts auditory features
- May be used standalone

# Two!Ears Auditory Model

Blackboard system



The screenshot shows the GitHub repository for the **TWOEARS / blackboard-system** project. The repository is on the **master** branch and contains 1,023 commits, 9 branches, 1 release, and 7 contributors. The repository is licensed under GPL-2.0. The commit history is as follows:

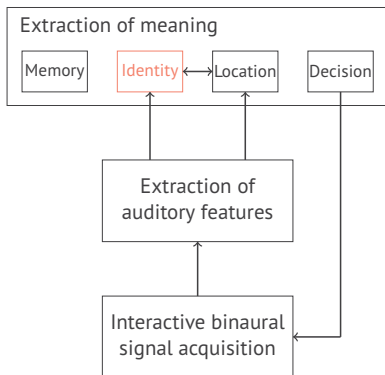
Commit	Author	Message	Time
1	Ning Ma	fixed angular distance bug	Latest commit: 17956a9 on Jan 19
2		doc: created evaluation folder for evaluation scripts	3 years ago
3		src: fixed angular distance bug	5 months ago
4		test: Remove further GMTK stuff	a year ago
5		gltignore: added GMTK binaries for MacOSX to third_party_software/gmtk-1.0.1.bin...	3 years ago
6		BlackboardSystem.xml: Remove Tools dependency	2 years ago
7		LICENSE: Add GOL2 license	2 years ago
8		README.md: Update doc links in README	6 months ago
9		star@BlackboardSystem.m: typo in star@BlackboardSystem.m	a year ago

- Knowledge sources for various tasks
- Scheduler



# Two!Ears Auditory Model

Training and testing pipeline



The screenshot shows the GitHub repository page for 'TWOEARS / Auditory-Machine-Learning-Training-and-Testing-Pipeline'. The repository has 1,075 commits, 6 branches, 2 releases, 8 contributors, and is licensed under BSD-2-Clause. The commit history is as follows:

Commit	Message	Time ago
ivo-4	bugfix in aggregateBlockAnnotations.m	13 hours ago
src	bugfix in aggregateBlockAnnotations.m	13 hours ago
test	mini fix in aggregateBlockAnnotations.m	14 hours ago
third_party_software	added third_party_software/extrema/LICENSE.txt	7 months ago
gitignore	number of sources test + fix for no parameters afe requests	10 months ago
AMLTP.xml	rename config	8 months ago
LICENSE	update of LICENSE and README.md (still immature)	7 months ago
README.md	added DOI citation	7 months ago
startAMLTP.m	rename and update start script	8 months ago

- Supports training
- Trained models → database

# Conclusions

## Software Development

Things that worked great:

- `git` and `github` excellent tools for collaborative software development
- `readthedocs` good place for creating documentation

Proposal for future projects:

- Make a workshop on the topic at the beginning
- Include a software engineer for complex software projects
- Try to avoid usage of closed-source software (e.g. Matlab)
- Training on reproducible research and publications

# Requirements on Data Management

- Data is collected and modified during the project:
  - HRIR/BRIR measurements for acoustic simulations
  - Training of machine learning stages
  - Listening test results
- All partners need seamless access
- Not all data can be made publicly available
- Potentially different versions of the same data set

⇒ Ideal solution: version control for data + rights management

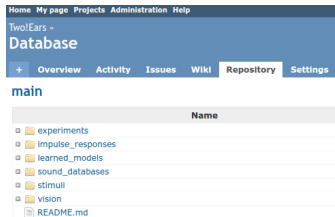
# Possible Approaches to Data Management

- `svn` works, but branching becomes buggy
- `git` may produce out of memory errors on the server
- `Git Large File Storage` was released during the project, but lacked a working server implementation
- Similar implementations from the community, like `git-media`, `git-annex`
- Commercial providers like `BitKeeper`

# Our Solution to Data Management

- Modified version of **git-media** for internal repository
- **svn** and **Redmine** for **public repository**
- In both cases you can download single files and subdirectories

## Web-frontend



## Matlab interface

```
fname = db.getFile('path/to/file');  
sig = audioread(fname);
```

# Conclusions

## Data management

Things that worked ok:

- `svn` + `Redmine` for `public data`
- `zenodo` for releasing single data sets

Proposal for future projects:

- Avoid complicated setups (like our `git-media`)
- Hope for better tools

# Summary

## Implementation of Open Science in Two!Ears

- Internal management of research data by **Redmine**
- Open source software & database, extensive documentation
- Includes full version history

## Some lessons learned...

- Open science requires training and qualification
- Seamless integration and usability of tools essential
- Counteracting evaluation measures

[github.com/twoears](https://github.com/twoears)

[github.com/spatialaudio](https://github.com/spatialaudio)